Chapter 4
Practical Aspects of Variogram Construction and Interpretation

4-1. General

a. Chapter 2 presented the mathematical foundation for geostatistics and the kriging technique. One theme that pervades the technique is the importance of the theoretical variogram. The theoretical variogram, or what we will often refer to simply as the variogram, is a mathematical function or model which is fitted to sample-variogram points obtained from data. Permissible models, which include those given in Chapter 2, belong to a family of smooth curves having particular mathematical properties and are each specified by a set of parameters. Chapter 4 will describe a sequence of stages for estimating and investigating sample variogram points and a calibration procedure for specifying the parameters of the variogram model eventually fitted to the sample points. Although the calibration procedure is largely an objective means for evaluating theoretical variograms, the process of obtaining sample variogram points and finalizing a theoretical variogram remains an art as much as a science. An understanding of the material presented in Chapter 2 as well as professional judgment achieved through experience in geostatistical studies is important in effectively using the guidelines presented in this section.

b. An accurate estimate of a variogram is needed from a kriging perspective because the correlation matrix used to obtain the kriging weights is constructed from the variogram values. Even more directly, the variogram affects the computation of the kriging variance (Equations 2-36 and 2-47) through the product of the kriging weights and variogram values. An accurate variogram also has utility outside the strict context of kriging. For example, in augmenting a spatial network with new data collection sites, the range parameter of the variogram could be used as the minimum distance of separation between the new sites and between new and existing sites to maximize overall additional regional information. In another non-kriging-specific application, the variogram is used in dispersion variance computations in which the variance of areal or block values is estimated from the variance of point-data values (e.g., Isaaks and Srivastava (1989), p. 480).

c. The stages of variogram construction are described using an example data set of ground-water elevations measured near Saratoga, WY (Lenfest 1986), that are summarized in Table 4-1 and whose relative locations are shown in Figure 4-1.

d. The sequence of steps in computing sample variogram points depends on the stationarity properties of the regional variable represented by the data. If the mean of the regional variable is the same for all locations, then it is said to be spatially

Table 4-1
Univariate Statistics for Example Data Sets[1]

Example Identifier	Number of Measurements	Transformation	Minimum (Base units)	Maximum (Base units)	Mean (Base units)	Median (Base units)	Standard Deviation (Base units)	Skewness (Dimensionless)
Saratoga		Drift	2,016.6	2,254.3	2,119.25	2,104.35	56.79	0.45
Water level A	83	Drift	25.6	65.68	42.30	38.54	10.13	1.03
Water level B	74	Drift	25.6	65.68	42.85	38.71	10.59	0.87
Bedrock A	108	None	22.64	80.48	44.42	42.82	10.76	0.89
Bedrock B	89	None	24.53	69.22	43.67	43.17	8.58	0.26
Water quality A	66	Natural log	2.08	8.01	5.19	5.59	1.75	-0.42

[1]Base unit for Saratoga, water levels, A and B, and Bedrock A and B is feet; base unit for water quality A is log concentration, concentration in micrograms per liter.

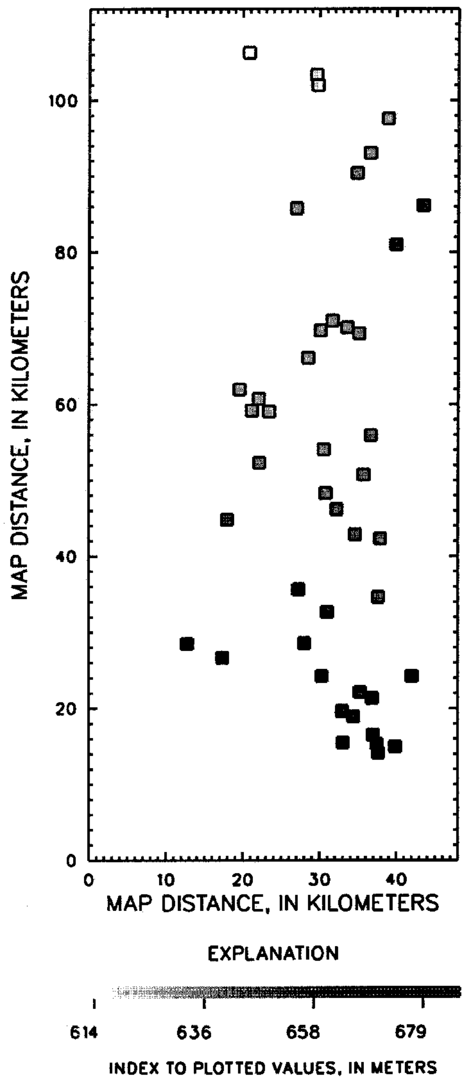


Figure 4-1. Measured water levels from Saratoga data

stationary; if the mean changes with location, then it is spatially nonstationary. Generally, if the data have a stationary spatial mean, the discussions in sections 4-3 and 4-7, which address nonstationarity and additional trend considerations, can be omitted. If the spatial mean is not stationary, as for this example data set, then sections 4-3 and 4-7 become important, and the sequence of stages for obtaining a variogram becomes an iterative procedure. All variogram and kriging computations for the Saratoga groundwater levels example were performed by the interactive kriging software described in Grundy and Miesch (1987).

4-2. General Computation of Empirical Variogram

a. As described in section 2-3, the variogram $\gamma(h)$ characterizes the spatial continuity of a regional variable for pairs of locations as a function of distance or lag h between the locations. This variogram is sometimes called the theoretical variogram because it is assigned a continuous functional form that expresses the spatial correlation for any lag in the region of analysis. The function is estimated by fitting one of the equations given in section 2-3 to empirical or sample variogram points $\hat{\gamma}(h)$ using data whose locations contribute only a finite number of lags. Although $\hat{\gamma}(h)$ characterizes the spatial correlation of the data, it is computed from residuals of the data off the spatial mean. Therefore, without prior knowledge of nonstationarity in the underlying spatial process, the first step in computing the sample variogram is to identify existing nonstationarity indicated for the spatial mean.

b. The approximation to Equation 2-19 begins by computing squared differences $D^2_{i,j}$ from the data values $z(\underline{x}_i), z(\underline{x}_2), \ldots z(\underline{x}_n)$ collected at locations $\underline{x}_1, \underline{x}_2, \ldots \underline{x}_n$

$$D^2_{i,j} = \left[z\ (\underline{x}_i) - z\ (\underline{x}_j) \right]^2 \qquad (4\text{-}1)$$

If the spatial mean is stationary, then the squared differences of the data are equivalent to the squared differences of the residuals, and sample variogram computations can be continued using the data themselves. If the spatial mean is strongly nonstationary, the plot of Equation 4-1 versus the distance between associated points may indicate a trend or drift that would need to be removed before further variogram computations could be made. Drift would have to be considered in HTRW studies, such as determining contaminant concentrations areally dispersed from localized sources or determining groundwater elevations following a local or regional gradient. In such studies sample variogram computations need to be made using residuals obtained by subtracting the estimated drift value at each location from the value of the datum at the location.

c. The data in Equation 4-1 are differenced without considering the relative direction between the locations; that is, $D^2_{i,j}$ is isotropically computed. A plot of $D^2_{i,j}$ versus $h_{i,j}$ for all i,j $(i>j)$, where $h_{i,j} = |\underline{x}_i - \underline{x}_j|$, produces a cloud of points whose properties govern the behavior of $\hat{\gamma}$. The central tendency of the cloud would generally increase with *h*. A substantial increase in the central tendency that persists for large h can indicate a nonstationary spatial mean. The cloud computed for the Saratoga data, with groundwater levels (*z*) in meters and distance (*h*) in kilometers, is shown in Figure 4-2 and does show increasing D^2 with increasing *h*, indicating potential non-stationarity.

d. Generally, there is a large amount of scatter in these plots, as seen in Figure 4-2, and this scatter can conceal the central behavior of D^2 with *h*. One way to estimate the central tendency and to minimize the effect of aberrant data values is to collect the D^2 into *K* bins or lag intervals of width $(\Delta h)_k$, $k=1,...K$ and assign to $\hat{\gamma}$ the average of the values of D^2 in each bin. This process is similar to the way data are placed in bins for obtaining histograms. The expression for the kth average bin value is

$$\hat{\gamma}(h_k) = \frac{1}{2N(h_k)} \sum D^2_{i,j} I_k(h_{i,j}) \tag{4-2}$$

where $N(h_k)$ is the number of squared differences that fall into bin *k*, and h_k is the lag distance associated with bin *k*. $I_k(h_{i,j})$ is an "indicator function" that has a value of one if the $h_{i,j}$ falls into bin *k* and zero otherwise (it only includes values of $D^2_{i,j}$ in the calculation that have an $h_{i,j}$ that falls into the bin). The lag value h_k can be the midpoint of the bin or it can be the average of the actual lag values for the points that fall in the bin.

e. To establish bins, either equal bin widths are specified and the distance between the two most separated data points, h_{max}, is subdivided according to these equal increments, or a *K* is chosen that defines the bin width. For the Saratoga data, a bin width of about 8 km established K=12 bins for γ. The $\hat{\gamma}$ points computed from the binned $D^2_{i,j}$ values of Figure 4-2 are shown in Figure 4-3. The lag plotting positions are the average *h* values in the bin. The symbol *x* indicates that $N(h)$ is less than 30 pairs for the particular bin and this differentiation will be discussed in section 4-3. Although the sample variogram is still preliminary, its general behavior at this stage is adequate to indicate if nonstationarity needs to be addressed before sample variogram refinement is undertaken.

4-3. Nonstationarity

a. An indication of substantial nonstationarity or drift in the spatial mean would be a parabolic shape through all lags in a plot of $\hat{\gamma}$. This shape occurs because differences between data contain differences in the drift component that increase as *h* increases. If Equation 2-16 is inserted into Equation 2-17, squaring the differences in μ greatly amplifies the increase with *h*. In these cases of drift, generally a low-order (less than three) polynomial drift in (*u*,*v*) is fitted to the data and subsequently subtracted from the data to obtain residuals. Trend surfaces are not necessarily limited to polynomial forms. For example, a numerical model of groundwater flow may be used to obtain residuals of groundwater head data.

b. In theory, the polynomial trend reflects a slowly varying drift in the spatial mean and, as such, one regional trend surface should be fitted to all the data. However, often the drift and residuals are obtained locally; that is, using moving neighborhoods of locations. Estimates of these values at any point are thus made using a reduced number (usually between 8 and 16) of surrounding locations. This is done because ultimately the kriging estimates are made using only the data values in the given neighborhood. Manipulating the kriging

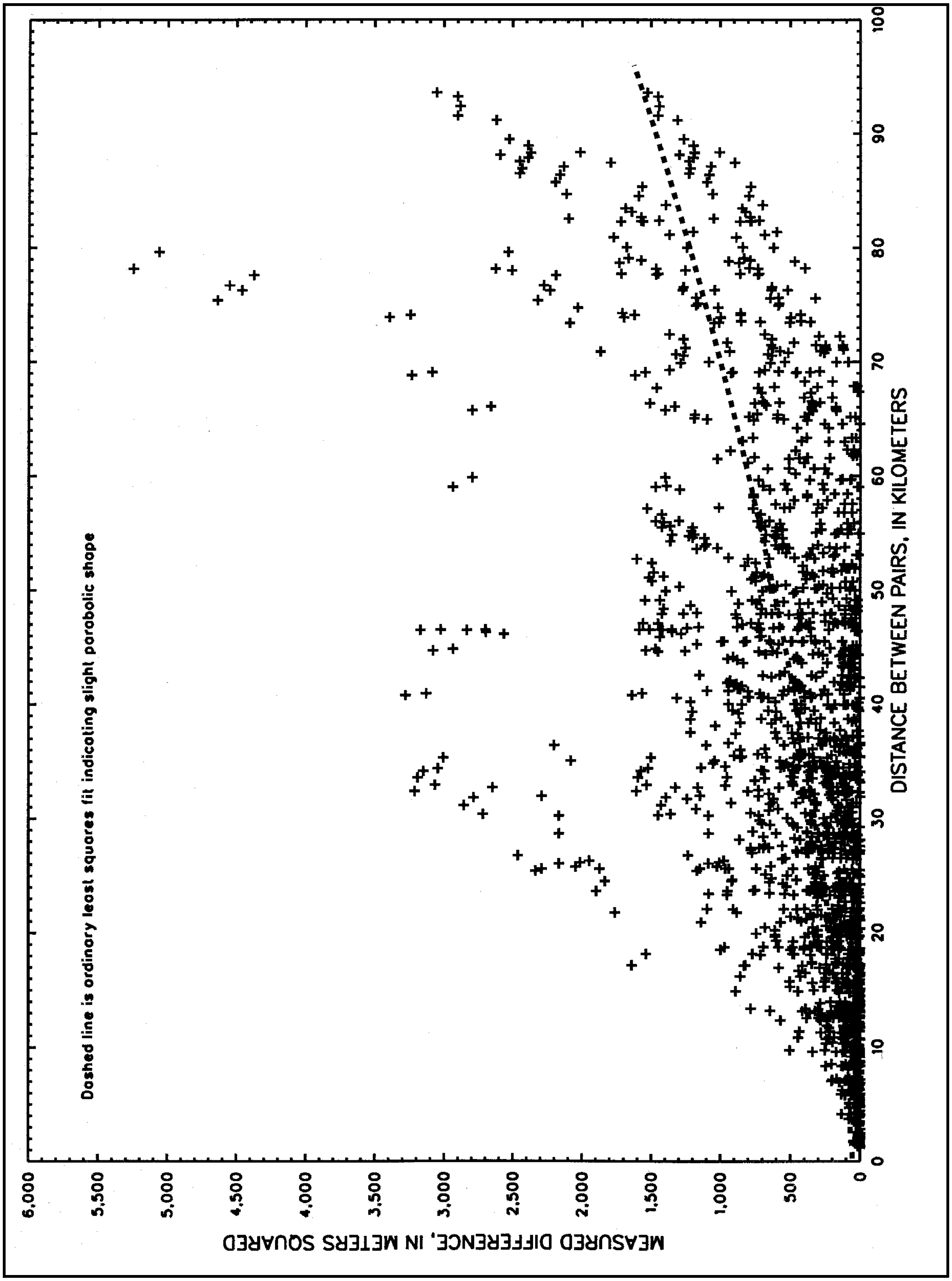


Figure 4-2. Squared differencing of values for all possible pairs of points for Saratoga data

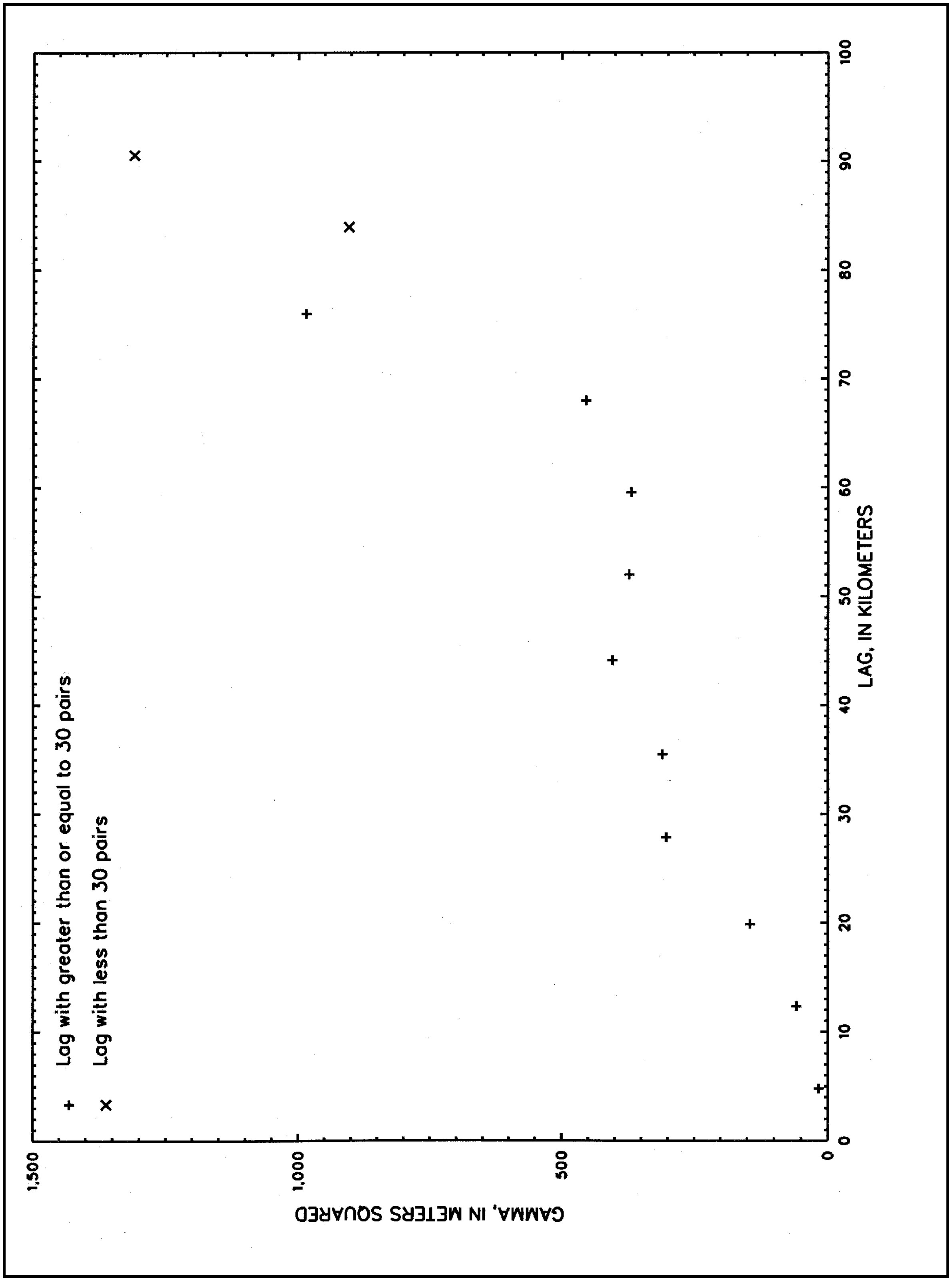


Figure 4-3. Initial sample variogram points for Saratoga data

matrices takes less time when a smaller number of data values are used to make estimates and these efficiencies can be significant when dealing with large data sets. Little accuracy is lost because the nearest neighbors are the most influential in the kriging weighting scheme.

c. A parabolic shape to $\hat{\gamma}$ for the Saratoga data is shown in Figure 4-3 for the sample variogram points plotted for lags up to about 32 km (the first four points) and for lags beyond about 56 km. The presence of a parabolic shape in the sample variogram points was not surprising, because examination of the data indicates a north-south gradient in the groundwater levels. The simplest polynomial trend, linear in u and v, was fitted to all the data using ordinary least-squares estimation. Residuals obtained by subtracting this regional trend surface from the data were used to reestimate $\hat{\gamma}$ in Equation 4-2 and the sample variogram for the residuals is shown in Figure 4-4.

4-4. Variogram Refinement

a. In the previous section, an initial $\hat{\gamma}$ was specified by points computed from Equation 4-2. In general, the larger $N(h_k)$ is for any bin or lag interval k, the more reliable will be the points defining $\hat{\gamma}(h_k)$. Also, the larger K is, the greater the number of sample variogram points shaping $\hat{\gamma}$. However, $N(h_k)$ and K are competing elements of $\hat{\gamma}$. Journel and Huijbregts (1978) suggest that each lag interval k should have $N(h_k)$ equal to at least 30 pairs. The American Society for Testing and Materials (Standard D5922-96) suggests 20 pairs for each lag interval. For small data sets the number of intervals may have to be small to guarantee either number of recommended pairs in all intervals.

b. It is difficult to determine the minimum number of data values n needed to satisfy the $N(h_k)$ requirements for all lag intervals of a sample variogram. Simple combinatorial analysis can establish a sample size needed to achieve a given total number of distinct pairs of items taken from the sample, but it does not address the spatial considerations needed for proper lagging. As an example, for data collected on a uniform grid and equal-sized bins, fixing an n to just satisfy the minimum $N(h_k)$ for the smaller lags will yield insufficient data pairs to meet the minimum $N(h_k)$ for the larger lags. Fixing an n to assure the minimum $N(h_k)$ for the larger lags will generally have $N(h_k)$ much greater than the minimum for the smaller lags. Therefore, the question of how much data is required to adequately compute a variogram should also address the relative locations of the data-collection sites.

c. The first 10 of the 12 bins for $\hat{\gamma}$ for the Saratoga data contained more than 30 data pairs. Therefore, the bin width can be decreased to get more points defining the early part of $\hat{\gamma}$. These bin-width adjustments can be made to refine $\hat{\gamma}$ whether it is computed from the data or from the residuals. A plot of $\hat{\gamma}$ for the residuals for the Saratoga groundwater elevations with the bin width narrowed to about 6.5 km is shown in Figure 4-5.

d. Spatial data are usually not collected on a uniform grid but occur in a pattern that reflects problem areas, accessibility, and general spatial coverage. In the Saratoga data set, nonuniform data spacing results in the number of data pairs in each bin, although still greater than 30, being highly variable among the bins. This variability yields different reliabilities for the points defining $\hat{\gamma}$. To establish a balance for $N(h_k)$ among the bins, variable bin sizes can be used so that each bin contains approximately the same (large) number of points. A bin with fewer points can be coalesced with an adjacent bin to form a wider bin with a greater number of points. Conversely, a bin with an excessive number of points can be subdivided into adjacent, narrower bins. The coalescing and subdividing procedure is largely trial and error, until the distribution of the pairs of points is satisfactory to the investigator.

e. The values of $\hat{\gamma}$ at the smaller lag values are the most critical to define the appropriate γ. Therefore, the trade-off between the number of bins and the number of data pairs within each bin can be varied for different regions of the sample

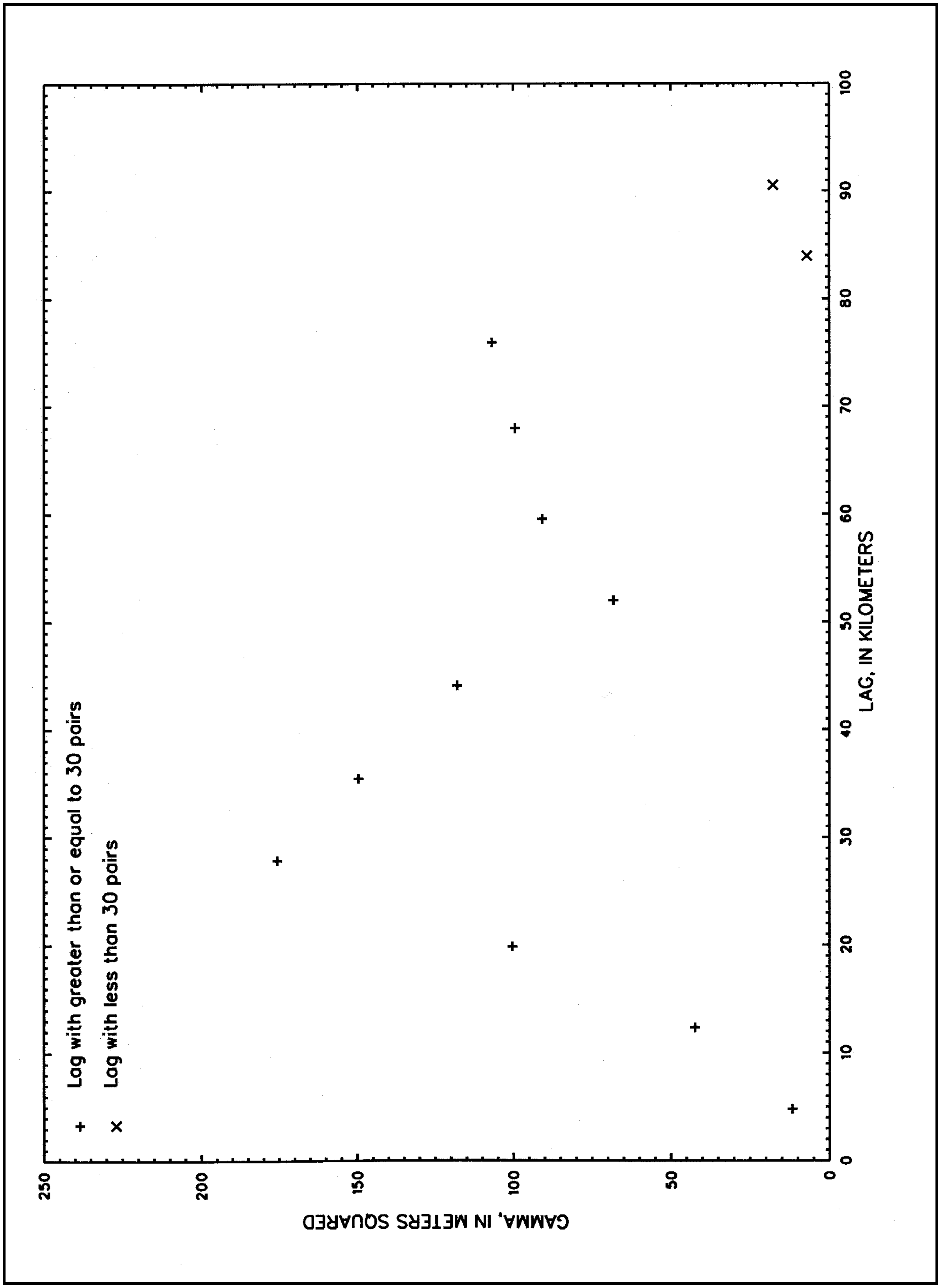


Figure 4-4. Sample variogram points for ordinary-least squares trend residuals of Saratoga data

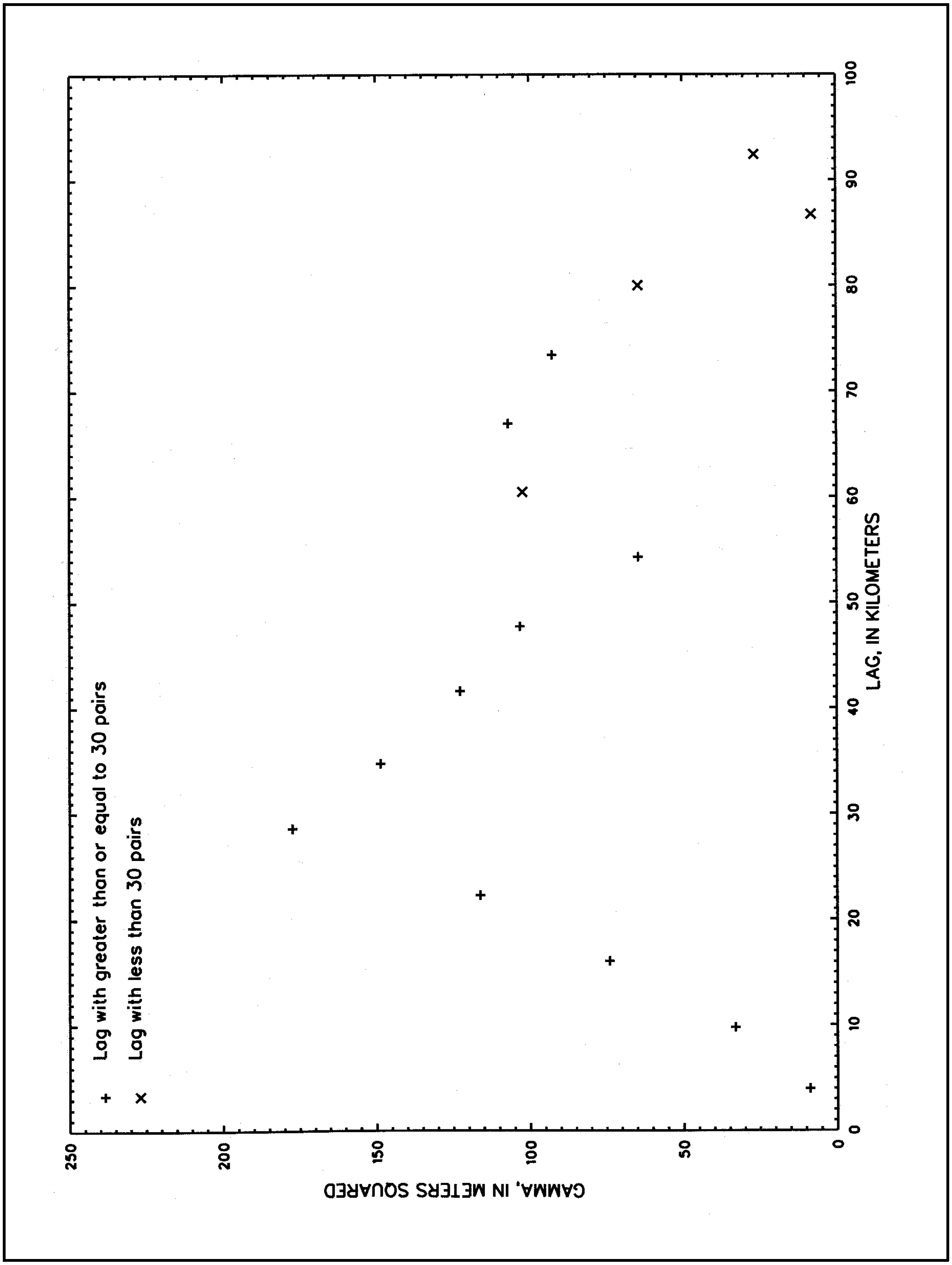


Figure 4-5. Sample variogram points for ordinary-least squares trend residuals of Saratoga data binned to 4 miles

variogram. At smaller lags, the numbers of data pairs per bin can be nearer the minimum $N(h_k)$ to define more bins. At larger lags, a smaller number of wider bins would be adequate. Knowing that the variogram should be a smooth function, ultimately the analyst visually decides when the sample variogram is sufficiently defined at all lags to adequately approximate a theoretical variogram.

4-5. Transformations and Anisotropy Considerations

a. Transformations.

(1) A transformation is applied to a data set generally for one of two interrelated purposes. First, a transformation can reduce the scale of variability of highly fluctuating data. This variability would especially occur with contaminant concentrations in which order of magnitude changes in data at proximate sites are not uncommon. The effects of such data would be erratic sample variogram points as exhibited by a large-amplitude, ill-defined sawtooth pattern of the lines connecting the points.

(2) Second, a proper transformation of data whose probability distribution is highly skewed often produces a set of values that is approximately normally distributed by mitigating the influence of problematic extreme data values. A data set with a normal distribution is important in kriging when confidence levels of the estimates are desired. This usage of confidence levels in a kriging analysis will be illustrated in Chapter 5.

(3) Among the more common transformations is the natural log transform. As an example, for this transformation, the $\hat{\gamma}$ will be the sample variogram values of logarithms, and subsequent kriged estimates will be logarithms. Another transformation that is often used, especially in spatial analyses of contaminant levels, is the indicator transformation described in Chapter 2. Although a transformation might achieve better-behaved sample variogram points, there are subtleties to consider in interpreting the kriging results of the transformed data or in back-transforming kriging results into the untransformed (original) units, as discussed in Chapter 1. If a satisfactory variogram of the original data cannot be achieved and a transformation is indicated, the sample variogram computation process must begin again with Equation 4-2. Even though no transformation was needed for the Saratoga data, an example using a logarithmic transformation and an example using the indicator transformation are presented in Chapter 5.

b. Directional variograms and anisotropy.

(1) Anisotropy in the data can be investigated by computing sample variograms for specific directions. Locations included in a given direction from any other location are contained in a sector of a circle of radius h_{max} centered on the location. The sector is specified by two angular inputs. The first is a bearing defining the specific direction of interest [measured counterclockwise from east ($=0^o$)] and the second is a (window) angle defining an arc of rotation swept in both directions from the bearing. Thus, in the terminology used here, the total angle defining a direction is equal to twice the window angle. Differences in sample variograms computed using these angle windows specified for different directions can be an indication of anisotropy.

(2) Anisotropy is generally either geometric or zonal. Geometric anisotropy is indicated by directional theoretical variograms that have a common sill value, but different ranges. The treatment of geometric anisotropy is dependent on the software used. The lags of the directional variograms can be scaled by the ratio of their ranges to the range of a standard or common variogram. In some cases, the lags of all directional variograms are scaled by their respective ranges, and a common variogram with a range of 1 is used. Groundwater contaminant plumes often have geometric anisotropy in which the prevailing plume direction would have a greater range than that of the transect of the plume.

(3) Zonal anisotropy is indicated by directional variograms that have the same range but different sills. Pure zonal anisotropy is usually not seen in practice; typically it is found in combination with geometric anisotropy. Such mixed anisotropy may be encountered if evaluating the variograms of three-dimensional HTRW sampling results. Variability of such data (as indicated by the sill of the variogram) may be significantly higher and the range significantly shorter in the vertical direction than in the horizontal direction. In order to model this mixture of anisotropic variograms, the overall variogram is set to a weighted sum of individual models of the directional variograms scaled by their ranges. In this process, called nesting, the choice of weights requires a trial and error approach with a constraint that the sum of the weights equals the sill of the overall variogram. The reader is referred to Isaaks and Srivastava (1989, pp. 377-390) for further information on both types of anisotropy.

(4) For a given number of data locations, directional sample variograms will necessarily have fewer points for any lag when compared to the points for the same lag in the omnidirectional variogram. Hence, there will be less reliability in the directional-variogram point values, which would be a critical constraining factor for small data sets or for a data pattern that does not conform to a direction of anisotropy. For a general idea of the sufficiency of the data to adequately determine any anisotropy, the computations of anisotropic sample variograms can be initially limited to two orthogonal directions with window angles of 45 deg.

(5) Directional sample variograms also can be used to further delineate nonstationarity of the spatial mean. If the omnidirectional sample variogram indicates a drift in the data, the directional variograms may determine the dimensionality of the drift. That is, although they may not establish the degree of the polynomial in the drift equation, the directional sample variograms can indicate the relative strengths of the drift in the u and v directions.

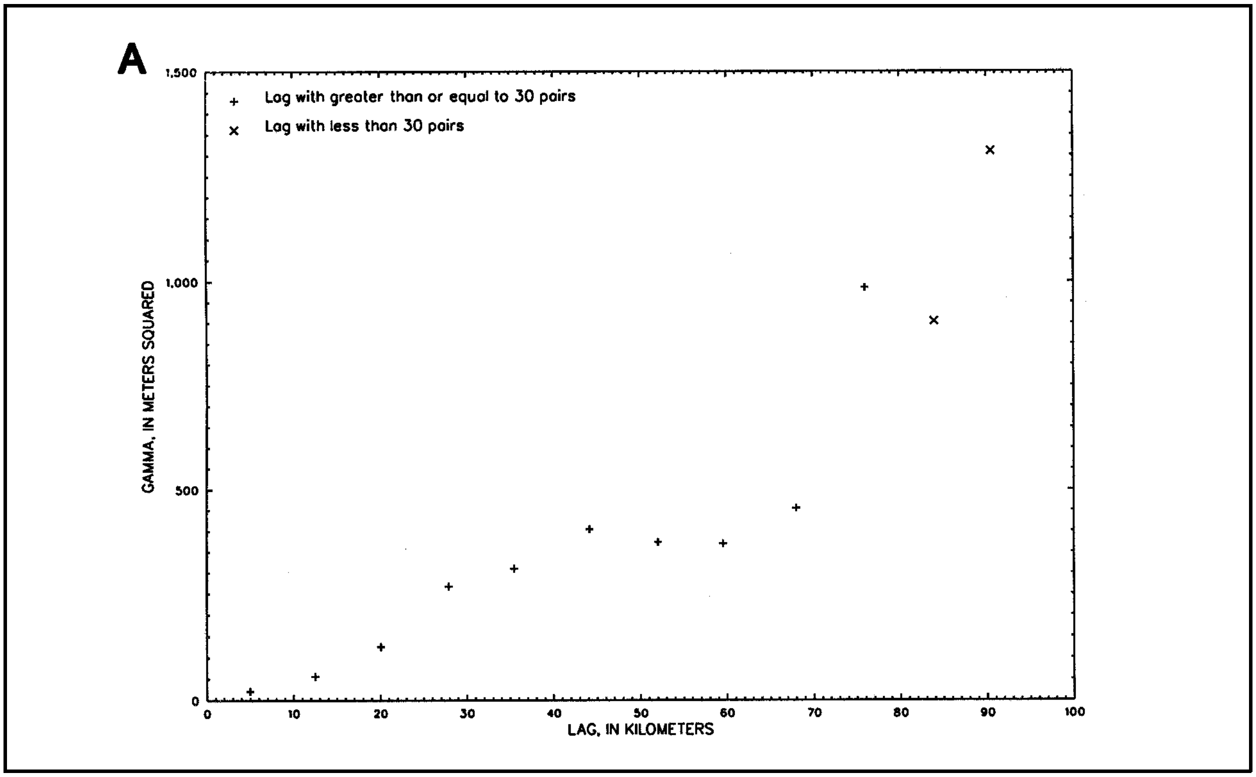
(6) The computed sample variograms for the general north-south and east-west directions for the Saratoga data are shown in Figure 4-6. The north-south variogram is specified by a direction angle of 90 deg and a window angle of 45 deg. The north-south variogram reveals the preferential north-south data alignment by mimicking the omnidirectional (direction angle = 0 deg and window angle = 90 deg) sample variogram of Figure 4-3. The east-west variogram is specified by a direction angle of 0 deg and a window angle of 45 deg. The lack of pairs of locations for the east-west variogram precludes a good analysis for this direction, but the overlap of the few sufficiently defined variogram points with the north-south variogram indicates a consistency of drift in the two directions. Because of this consistency, an isotropic variogram is assumed for the Saratoga residuals. An example of anisotropic variograms is described in Chapter 5.

4-6. Fitting a Theoretical Variogram to the Sample Variogram Points

a. General.

(1) The importance of adequately defining the bin values of a sample variogram is substantiated by the need to accurately generalize the data-based behavior of the sample variogram by a theoretical variogram γ. The parameters controlling the specific behavior of theoretical variograms are the nugget value, the range, the sill, or in the case of a linear variogram, a slope parameter. Of these parameters, the nugget and the sill can be related to properties and statistics of the data.

(2) The nugget is essentially the extrapolation of the sample variogram to a lag of zero. It reflects the uncertainty of the variogram at lags that are much smaller than the minimum separation between any two data locations. The nugget value can include measurement error variance, and an estimate of this variance will approximate a minimum value of the extrapolation.



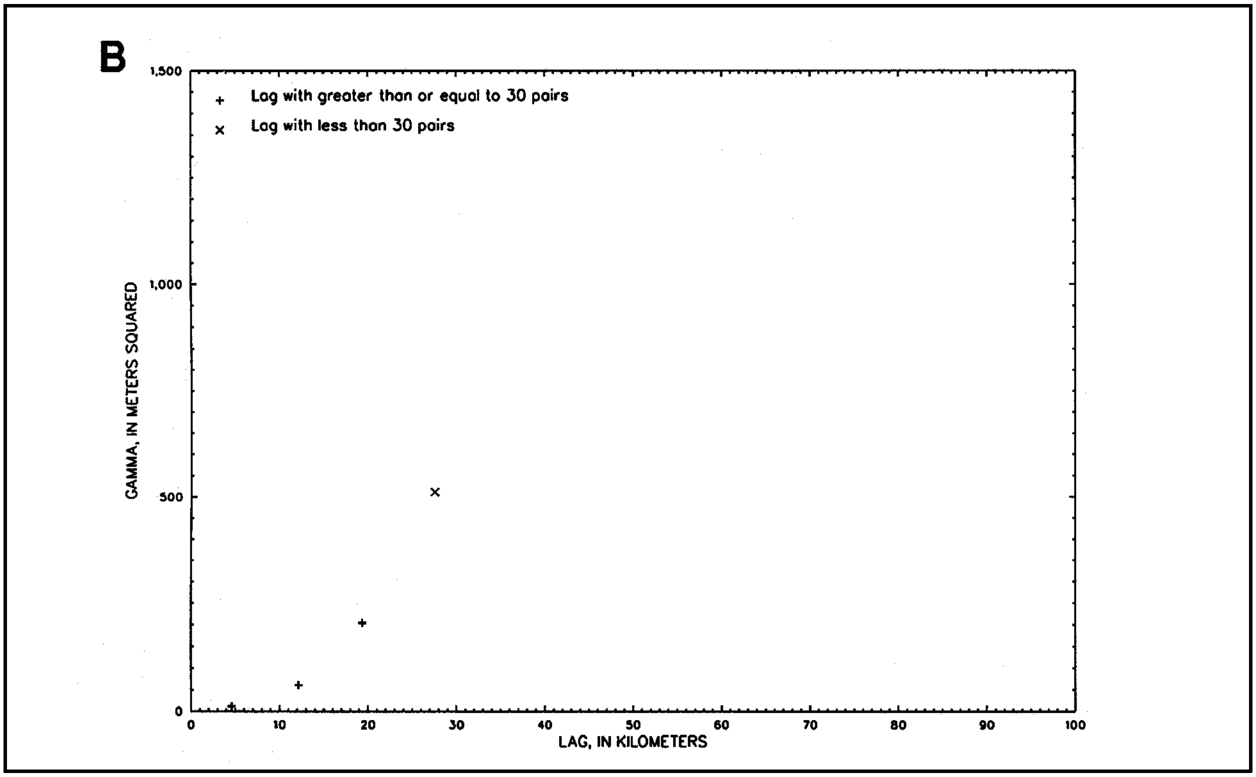


Figure 4-6. Initial directional sample variogram points for raw Saratoga data--A, north-south and B, east-west

(3) The sill determines the maximum value of a variogram and approximates the variance of the data. However, the points defining $\hat{\gamma}$ take precedence over the sample variance in locating the sill. Some variograms are unbounded, and others may only reach a sill value asymptotically. A defined sill allows conversion of the variogram to a covariance function using Equation 2-27, which is generally done because computations in the kriging algorithms are more efficiently performed using a covariance function.

(4) Fitting a function to the sample variogram values can range from a visual fit to a sophisticated statistical fit. A statistical fit is an objective method as long as the choice of bins and weighting of the sample variogram points remain fixed. However, because the inputs will vary with investigators, inherent subjectivity persists as in a visual fit. A final calibration of the variogram parameters would be based on the kriging algorithm and, thus, either of the initial fitting methods at this stage would suffice.

(5) Because the initial part of the variogram has the most effect on subsequent kriging output, a good estimate of the nugget value becomes a most important first step. The range and the sill, in that order, complete the ranking of the influence of variogram parameters on the output of a geostatistical analysis. Whatever the fitting method used, the theoretical variogram needs to be supported by the sample variogram values. For variograms with a range parameter, this support should extend to the range. Journel and Huijbregts (1978) suggest that this support should be through one-half the dimension of the field or essentially through one-half the maximum lag distance of the sample data.

(6) Most geostatistical studies can be successfully completed using the following four singular theoretical variogram forms: exponential, spherical, Gaussian, and linear functions (Figure 2-3). For the example variogram determination described in this section, only one of these singular forms will be selected; however, positive linear combinations of these forms also are acceptable as theoretical variograms (see section 4-5*b*). Geometric relationships to aid in obtaining parameters for the four variogram forms are described in the following sections and are illustrated in Figure 2-3 for reference.

b. Exponential variogram.

The exponential variogram (Equation 2-23) is specified by the nugget *g*, sill *s*, and a practical range value *r*. The range is qualified as practical because the sill is reached only asymptotically. The initial behavior of the exponential variogram is different from the behavior of the spherical variogram in that the convex behavior extends to the nugget value (Figure 2-3). Again, a nugget value and a sill value are first specified based on the $\hat{\gamma}$ points. The practical range is chosen so that the value of the resulting exponential function evaluated at the practical range lag is 95 percent of the sill value. The specified exponential function would mesh with the sample variogram points at least through the practical range lag. An initial estimate of the practical range can be made by checking if the intersection of the sill value with a line tangent to the variogram at the nugget is at a lag value equal to one-third of the assumed practical range value as illustrated in Figure 2-3. Examples of the exponential variogram may be found in spatial studies of sulfate and total alkalinity in groundwater systems (Myers et al. 1980).

c. Spherical variogram. The spherical variogram parameters (Equation 2-24) are a nugget value *g*, a range *r*, and a sill *s*. At smaller lag values the sample variogram points indicate linear behavior from the nugget that then becomes convex and reaches a sill value at some finite lag (Figure 2-3). A sill is estimated, and a line drawn through the points of the initial linear part of the variogram would intersect the sill at a lag value approximately equal to two-thirds of the range. With these estimates of the parameters, a spherical variogram is defined that should be supported by the sample variogram points. If the spherical plot does not fall near the sample variogram points, adjustments need to be made to the parameter estimates and the subsequent fit evaluated. Although the spherical variogram is one of the most often used models for real valued spatial studies, it seems to be a predominant model for indicator values at various cutoff levels as, for example, in a study of lead contamination (Journel 1993).

d. *Gaussian variogram*. The Gaussian variogram parameters (Equation 2-25) are a nugget value *g*, and a sill *s*, and this variogram also has a practical range *r*. The Gaussian variogram is horizontal from the nugget, becomes a concave upward function at small lags, inflects to concave downward, and asymptotically approaches a sill value (Figure 2-3). After a nugget value and sill value are specified based on the points, the variogram value at a lag of one-half the estimated practical range will be two-thirds of the sill value. Again, this fitted variogram needs to be supported by the $\hat{\gamma}$ points to a reasonable degree. As will be described in the example using the Saratoga data, the Gaussian variogram often is used where the variable analyzed is spatially very continuous, such as a groundwater potentiometric surface.

e. *Linear variogram*. Parameters for a linear variogram (Equation 2-26) are a nugget value *g*, and a slope *b*. Sample points indicating a linear variogram would increase linearly from the nugget value and fail to reach a sill even for large lags (Figure 2-3). With the nugget as the intercept, the slope is computed for the line passing through the $\hat{\gamma}$ points. A pseudosill *s* can be defined as the value of the line at the greatest lag, h_{max}, between any two locations. This lag becomes the defacto range *r* for a linear variogram. Examples of the usage of the linear variogram occur in hydrogeochemical studies of specific conductance and in studies of trace elements such as barium and boron (Myers et al. 1980).

4-7. Additional Trend Considerations

a. If a drift in the data is indicated as in section 4-3, the theoretical variogram of residuals that has been fitted thus far is used to update the drift equation. Although ordinary least squares often suffices for computing a polynomial drift equation, drift determination itself is a function of γ when the data are spatially correlated. But γ cannot be estimated until a drift equation is obtained to yield the residuals. Therefore, obtaining a sample variogram and a subsequent theoretical variogram from drift residuals of a specified drift form is an iterative process (David (1977), pp. 273-274) framed by the following steps:

(1) An initial variogram is specified and drift coefficients are computed to obtain residuals. For this step, a pure nugget (i.e. constant) variogram can be used to compute the initial estimates of the drift coefficients. This is an ordinary least-squares estimate of the drift yielding a first-iteration sample variogram of residuals.

(2) A theoretical variogram is fitted to the sample variogram of the residuals and is used to obtain updated drift coefficients.

(3) The residuals from the drift obtained in step b are used to compute an updated sample variogram.

(4) The sample variogram computed at the end of step 3 is compared to the sample variogram of step 2. If the two sample variograms compare favorably, then the theoretical variogram from step 2 is accepted as the variogram of residuals for subsequent kriging computations. If the sample variogram from step 3 differs markedly from the sample variogram of step 2, steps 2-4 are repeated using the sample variogram of the most recent step c.

b. Generally, the plot of the points of $\hat{\gamma}$ from a set of residuals will initially increase with *h*, reach a maximum, and then decrease as seen in Figure 4-4. This typical haystack-type behavior, discussed by David (1977, pp. 272-273), is attributed to a bias resulting from the estimation error in the drift form and its coefficients. Thus, this behavior in the variogram of the residuals generally would more readily occur with a higher degree of drift polynomial. This behavior should not prohibit acceptable variogram determination because the initial points of the sample variogram of residuals are still indicative of the theoretical variogram. For example, the lag associated with the maximum of $\hat{\gamma}$ of the residuals can be a good first approximation for the range of the theoretical variogram.

4-8. Outlier Detection

a. Outliers in a data set can have a substantial adverse effect on $\hat{\gamma}$. However, divergent data values can be screened for evaluation using a Hawkins statistic (Hawkins 1980), which is described in the context of kriging by Krige and Magri (1982). A neighborhood containing 4 to 10 data points, approximately normally distributed, around each suspected outlier must be defined. Despite potential outliers in the data set, a best guess initial theoretical variogram also is needed.

b. The Hawkins statistic is obtained by comparing a suspect datum to the mean value of the 4 to 10 surrounding data, the smaller number being sufficient if the variability is lower. The spacing between these surrounding points is accounted for by the properties of the chosen variogram. A value for the statistic of 3.84 or higher would indicate an outlier on the basis of a 95-percent confidence interval. A larger number of surrounding points has the direct effect of increasing the magnitude of the statistic. Anomalous points are removed from the data set and the procedures described for obtaining the sample variogram are repeated for the smaller data set. There were no outlier problems in the Saratoga data.

c. There is debate among geostatisticians regarding the merit of automated outlier-detection methods. A procedure such as that described here is presented as an investigative tool with the understanding that the investigator will also use attendant justification along with a Hawkins-type statistic to ultimately decide if a data value is discarded as a true outlier or retained as a valid observation. In some situations, highly problematic data values are removed for computation of the sample variogram points but are reinstated for kriging.

4-9. Cross-Validation for Model Verification

a. General.

(1) Parameters of the theoretical variogram obtained from the initial fitting and refinement of the sample variogram are calibrated using a kriging cross-validation technique. In this procedure, the fitted theoretical variogram is used in a kriging analysis in which data values are individually suppressed and estimates made at the location using subsets of the remaining points. As described in section 4-3, these subsets are the data points in a moving neighborhood surrounding the point under consideration. The calibration estimate made at each data location requires a matrix inversion, which could be very time-consuming if all remaining data locations were used to construct the matrices rather than just those within a neighborhood of a limited search radius.

(2) After kriged values at all data locations have been estimated in the above manner, the data are used with their kriged values and kriging standard deviation to obtain cross-validation statistics. A successful calibration is based on criteria for these statistics, which are described in the next section. If the criteria cannot be reasonably met by adjusting the parameters in the given theoretical variogram function, then calibration should be reinitialized with a different theoretical variogram function. In some data sets with nonstationary spatial means, the drift polynomial may have to be changed as well as the variogram to achieve a satisfactory calibration.

b. Calibration statistics.

(1) The kriging cross-validation error e_i corresponding to measurement $z(\underline{x}_i)$ is defined as

$$e_i = z(\underline{x}_i) - \hat{z}(\underline{x}_i) \tag{4-3}$$

where $\hat{z}(\underline{x}_i)$ is the kriged estimate of $z(\underline{x}_i)$ based on the remaining n-1 measurements in the data set.

The kriged estimate is obtained by ordinary kriging if the spatial mean is constant or by universal kriging if the spatial mean is not stationary. A reasonable criterion for selecting a theoretical variogram would be to minimize the squared errors, $\sum e_i^2$, with respect to the variogram parameters. However, unlike ordinary least-squares regression, which also minimizes the sum of squared errors, simply minimizing the squared errors is not sufficient for kriging because the resulting model can yield highly biased estimates of the kriging variances, $\sigma_k^2(\underline{x}_i)$, where $\sigma_k^2(\underline{x}_i)$ is the kriging variance at location $\underline{x}_i$. This simple minimization would give unrealistic measures of the accuracy of the kriging estimates. To guard against such bias, an expression for the square of a reduced kriging error is defined:

$$\tilde{e}_i^2 = \frac{e_i^2}{\sigma_K^2(\underline{x}_i)} \tag{4-4}$$

where the kriging variances are computed using either Equation 2-36 or 2-47. If the kriging variance is an unbiased estimate of the true mean-squared error of estimate, then the reduced kriging errors would have an average near one. Therefore, the standard cross-validation procedure for evaluating a theoretical variogram is:

$$\min \left(\frac{1}{n} \sum_{i=1}^{n} e_i^2 \right)^{0.5}$$
$$\text{subject to} \left(\frac{1}{n} \sum_{i=1}^{n} \tilde{e}_i^2 \right)^{0.5} \approx 1 \tag{4-5}$$

(2) The expression to be minimized is called the kriging root-mean-squared error and the constraint is called the reduced root-mean-squared error. The reduced root-mean-squared error should be well within the interval having endpoints $1 + \left(2\sqrt{\frac{2}{n}}\right)$ and $1 - \left(2\sqrt{\frac{2}{n}}\right)$ (Delhomme 1978). An additional check on the cross-validation results is the unbiasedness condition where $\frac{1}{n}\sum e_i \approx 0$.

(3) As indicated in Chapter 2, if probabilistic statements concerning an actual value of Z at an unmeasured location are to be made relative to the kriged estimate and the kriging variance at the location, it is necessary to explore the distribution of the cross-validation kriging errors. In particular, it is desirable that the reduced errors, $\tilde{e}_i$ =1,2...,n, are approximately normally distributed with mean 0 and variance 1. A histogram or normal probability plot of the reduced kriging errors can be used to assess the validity of assuming a standard normal distribution for the reduced kriging errors. Additionally, if the distribution of reduced kriging errors can be assumed to be standard normal, outliers not detected using the method discussed in section 4-7 may be detected by comparing the absolute values of the reduced kriging errors to quantiles of the standard normal distribution.

(4) Using the Saratoga data, a spherical variogram was fitted to the refined sample variogram of the residuals. The estimated nugget was about 1.49 m^2, the sill was 133.8 m^2, and the range was about 48 km. Because of difficulty in determining an exact extrapolated value for the nugget, the value of 1.49 m^2 was selected based on an estimated measurement error related to obtaining water levels at the well depths in the Saratoga valley.

(5) After two iterations using drift residuals, as described in section 4-7, a final variogram was chosen with a nugget of 1.49 m^2, a sill of 148.6 m^2, and a range of 44.8 km (Figure 4-7). These parameters defined the theoretical variogram used to obtain the cross-validation errors using universal kriging with an assumed linear drift. The best combination of statistics that could be obtained after several attempts at refining the model were a root-mean-squared error of 3.45 m and a reduced root-mean-squared error of 0.5794. The

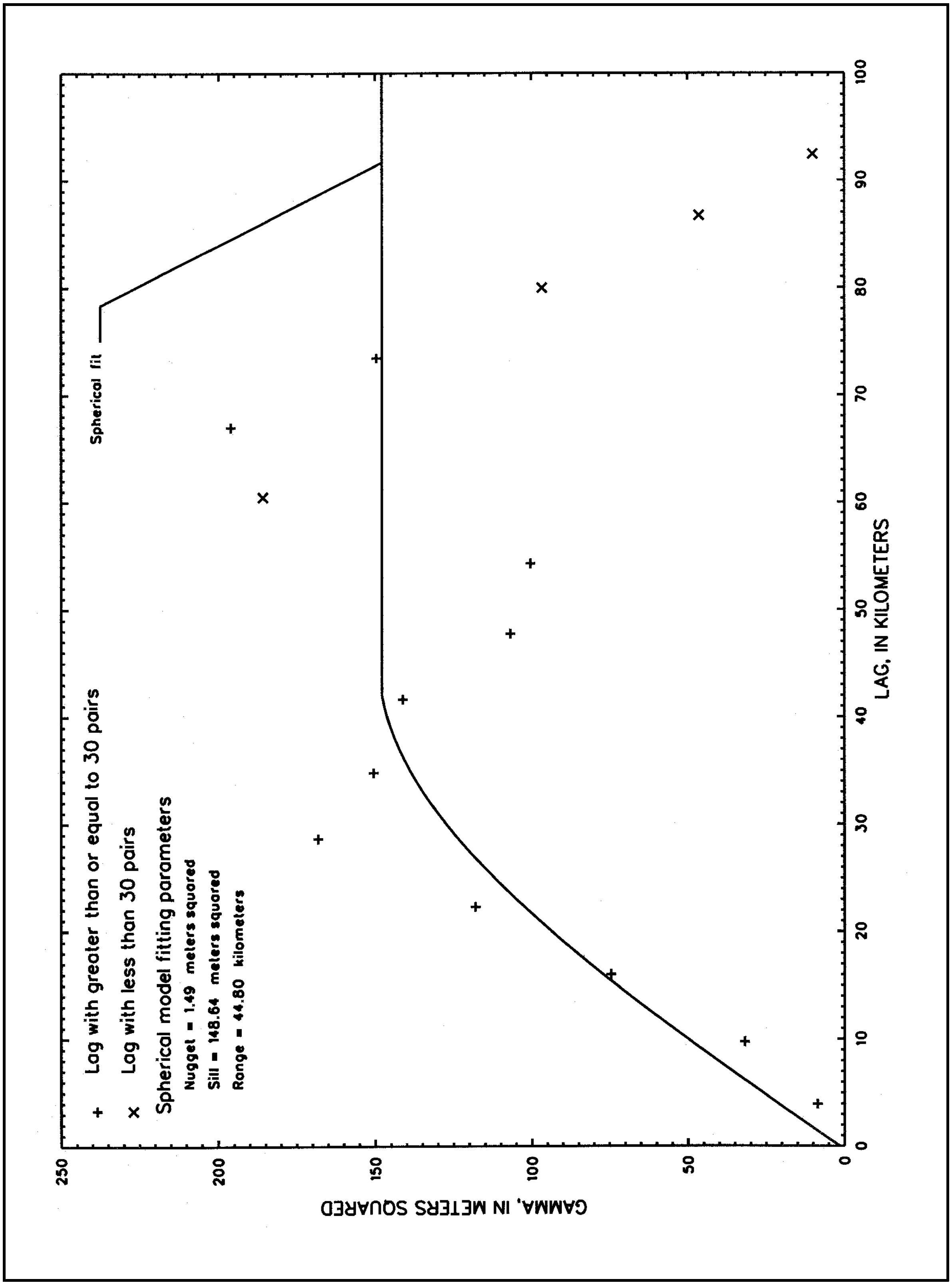


Figure 4-7. Sample variogram points and theoretical spherical fit for iterated Saratoga residuals

reduced-root-mean-squared error is too small, indicating that the kriging variances produced by the model are too large compared to the actual squared errors. This fact, coupled with the rather large root-mean-squared error, makes the theoretical variogram model unacceptable. In section 4-9*c*, a Gaussian variogram is fitted to the data that produces much better cross-validation results than the results for the spherical variogram.

c. Variogram-parameter adjustments.

(1) If any of the cross-validation statistics vary unacceptably from their suggested values, minor adjustments to the variogram parameters can be made to attempt to improve the statistics. Whatever modifications are made to the parameters, they should not have to be so severe that the variogram function drastically deviates from the sample variogram points. If the support of the sample variogram points is compromised in order to achieve acceptable cross-validation results with the given drift-variogram model, a different drift-variogram combination should be investigated.

(2) A reduced root-mean-squared error that is unacceptable may be improved upon by adjusting the range parameter or the nugget value of the variogram. Modifying the range parameter should be considered first and any shifts in the nugget value should be minimal and made only as a final recourse. Calibration errors are relatively insensitive to minor adjustments of the sill.

(3) If the reduced root-mean-squared error is too small, as in the Saratoga example, extending the range (equivalent to decreasing the slope for a linear variogram) will decrease the kriging variance and thus increase the reduced root-mean-squared error. If a shift in the nugget value is required, a decrease in the nugget will reduce the kriging variance. If the reduced root-mean-squared error is too large, then a contraction of the range or a positive shift in the nugget value can be made, keeping in mind the above caveat of priority and extent of the changes. Changes in these parameters, generally, also will have an effect on the mean-squared error. The larger the nugget is as a percentage of the sill, the larger the mean-squared error will be. In general, improvements in one statistic are usually made at the expense of the other statistics. The optimization of the statistics as a set is, in effect, a trial and error procedure that is operationally convergent.

(4) Reduced kriging errors may not approximate a standard normal distribution. If this is the case, a transformation of the data may be needed to achieve a more normal distribution, and the variogram estimation procedure would be repeated.

(5) Because no convergence could be reached for parameter values of a spherical variogram for the Saratoga data, a Gaussian theoretical variogram was fitted to the sample variogram of residuals in Figure 4-4. This choice was made because the initial sample variogram points could be interpreted to have a slight upward concavity, but eventually reached a sill. This behavior can be attributed to correlation rather than to further drift. After an iterated cross-validation with the Gaussian parameters, a Gaussian variogram with a nugget of 1.49 m^2, a sill of 185.81 m^2, and a range of 27.52 km (Figure 4-8) yielded a root-mean-squared error of 2.33 m and a reduced-root-mean-squared error of 1.083. The mean cross-validation error is 0.0195 m. These values represent an improvement over the spherical variogram and were deemed acceptable for the Gaussian variogram.

(6) A probability plot of the reduced kriging errors using the final Gaussian variogram is shown in Figure 4-9. It is reasonably linear between two standard deviations and, thus, approximates a standard-normal-distribution function. Finally, a plot in Figure 4-10 of the data versus their kriged estimates indicates that the linear drift-Gaussian variogram model selected for the Saratoga data would produce accurate estimates of groundwater elevations for interpolation or contour gridding in the region.

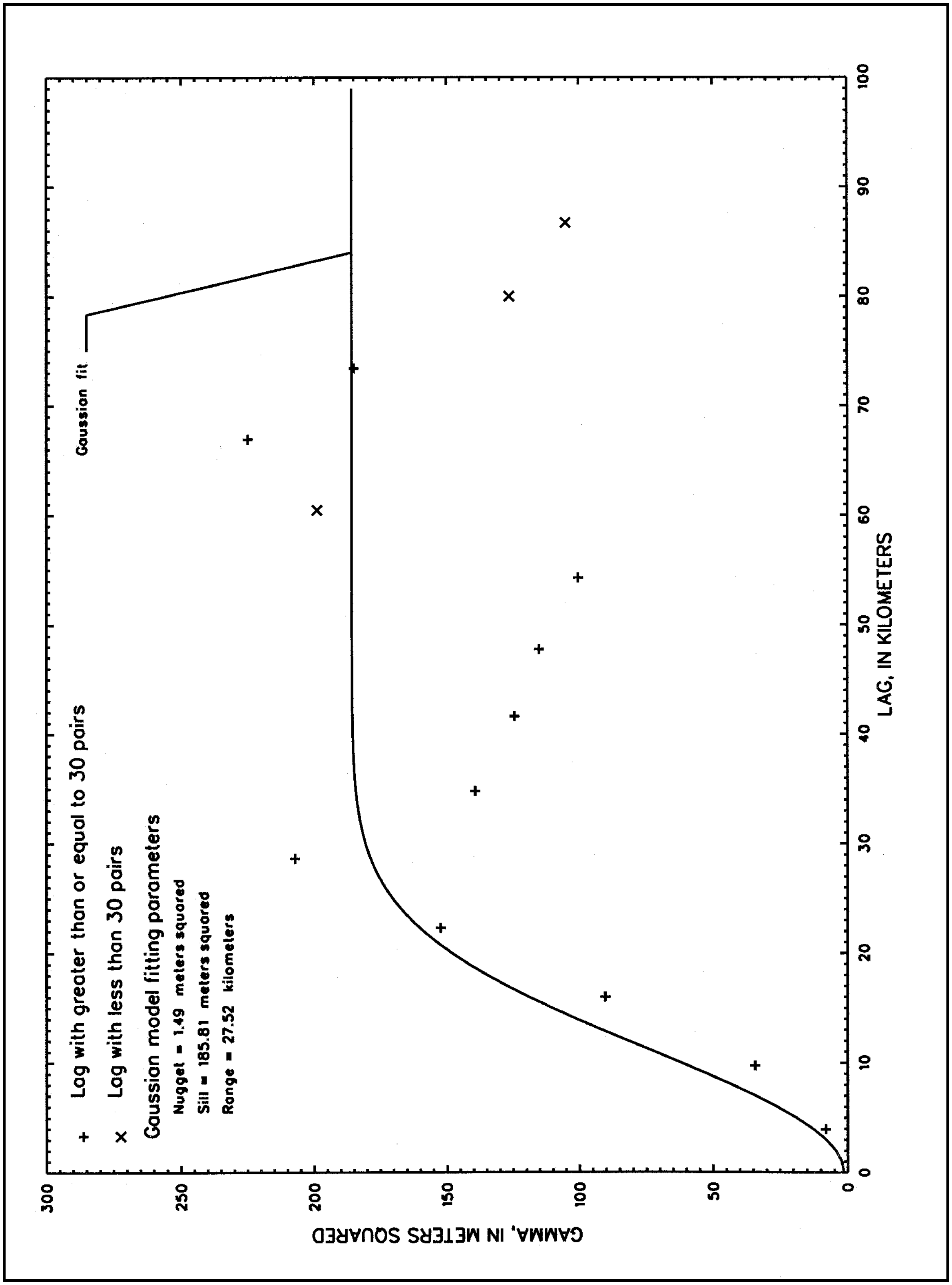


Figure 4-8. Sample variogram points and theoretical Gaussian fit for iterated Saratoga residuals

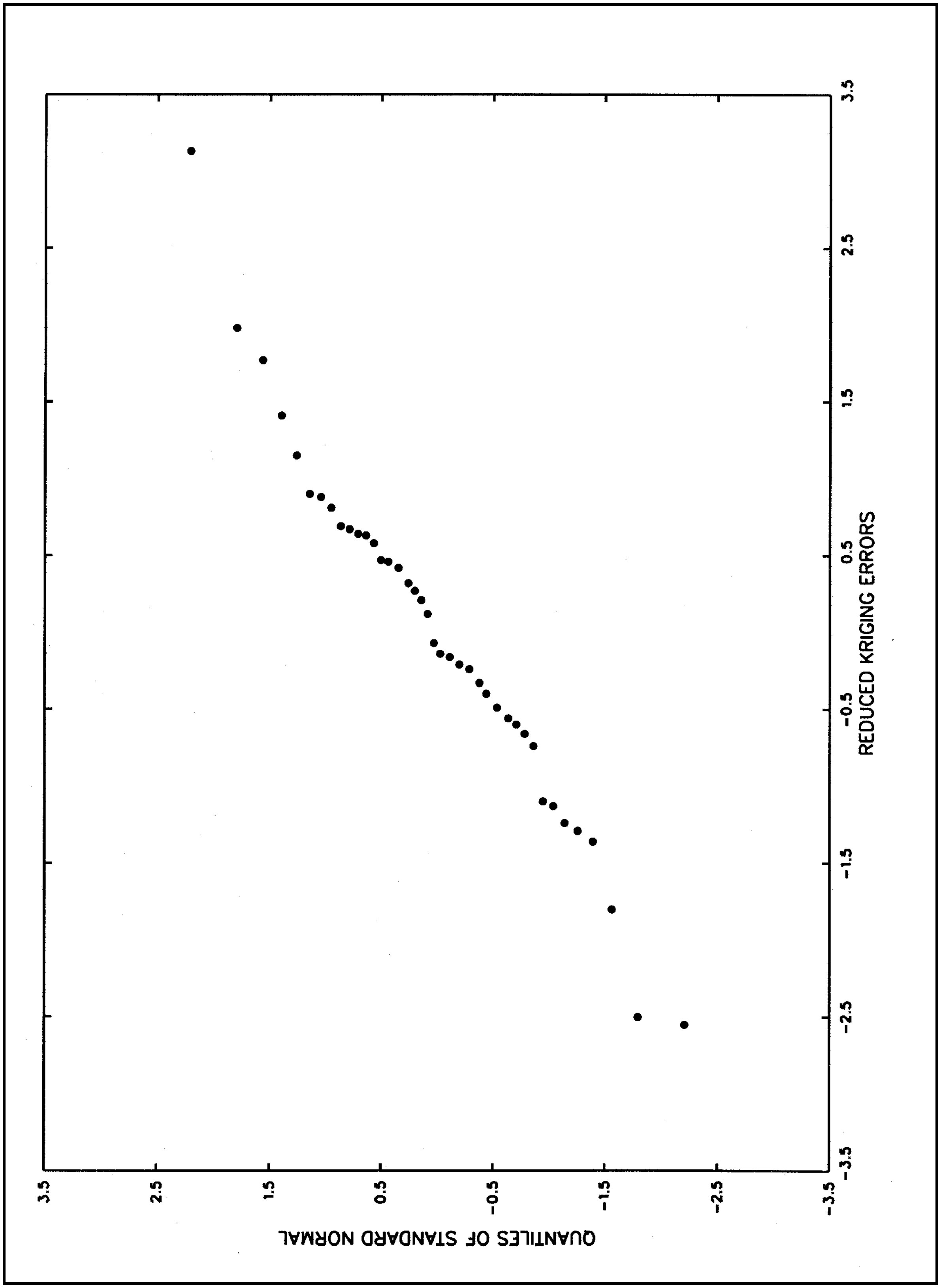


Figure 4-9. Cross-validation probability plot for Saratoga data

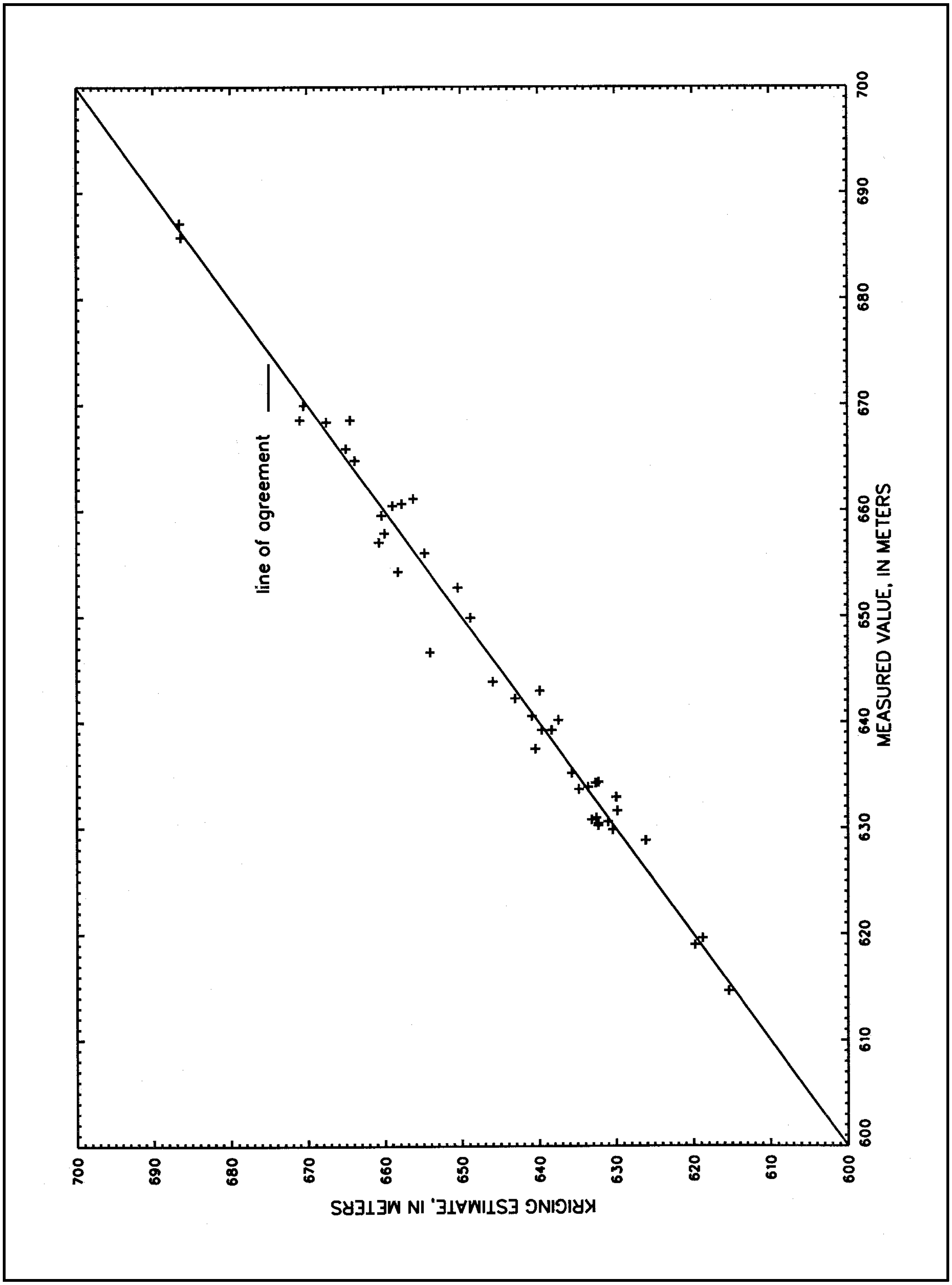


Figure 4-10. Scatter plot of measured versus kriging estimates from cross-validation of Saratoga data